Learning outcomes

✓ Why is Data Literacy Essential?	✓ Impact of Data Literacy
✓ How to become Data Literate?	✓ Data Literacy Process Framework
✓ Data Security & Privacy	✓ How is Data Security & Data Privacy related to AI?
✓ Acquiring Data, Processing & Interpreting Data	✓ Data Acquisition / Acquiring Data
✓ Usability, Features & Processing of Data	✓ Data terminology
✓ Methods of Data Interpretation	✓ Types of Data Interpretation
✓ Importance of Data Interpretation	✓ Using Tableau for Data Presentation

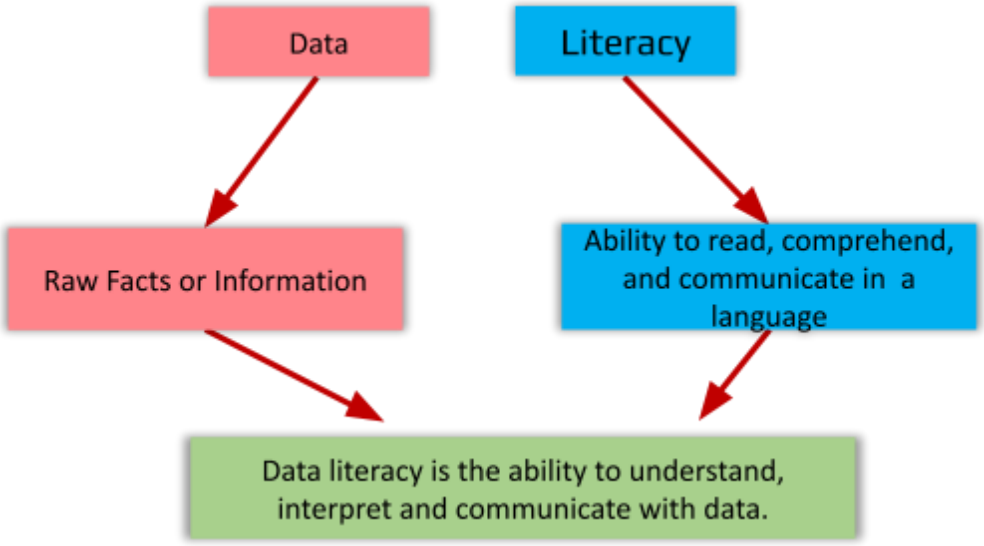
Data refers to any collection of raw facts, figures, or statistics that can be stored and processed by a computer. It can be in different forms like numbers, text, images, audio, and video etc.

"**Literacy** refers to the ability to read, comprehend and use information effectively.

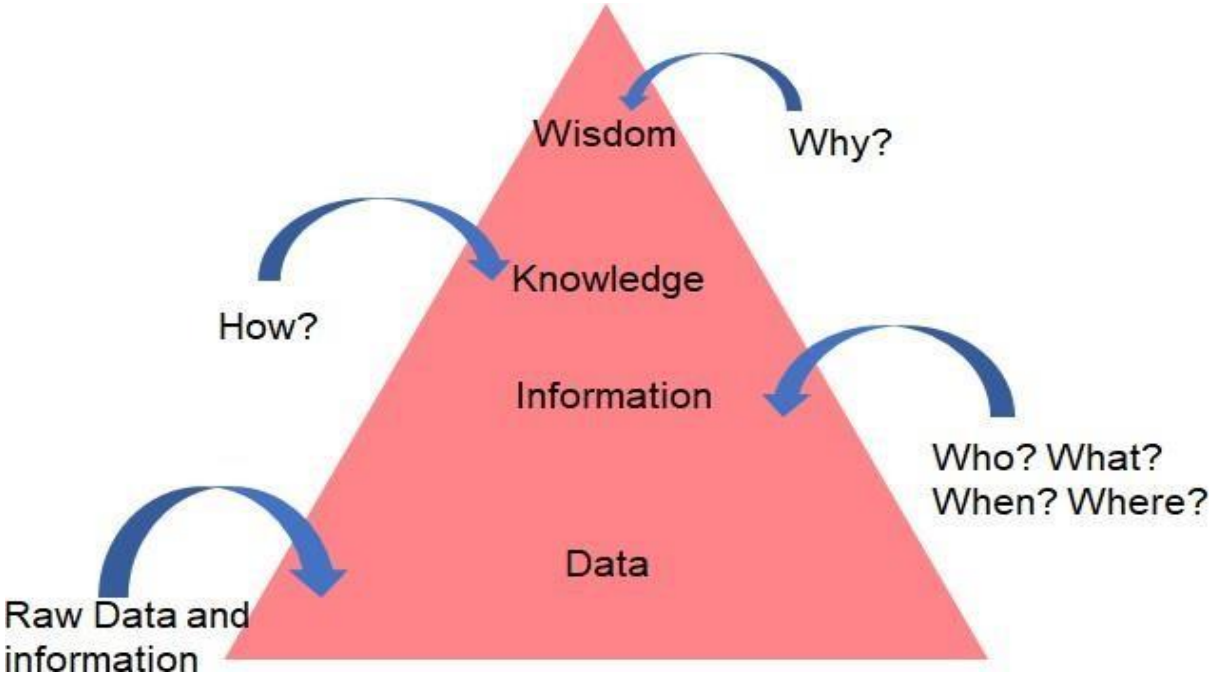
Data + Literacy = Data Literacy

Data literacy means knowing how to understand, work with, and talks about data.It's about being able to collect, analyse, and show data in ways that make sense.

"Data literacy is essential because it enables individuals to make informed decisions, think critically, solve problems, and innovate."



Data Pyramid and Its Different Stages



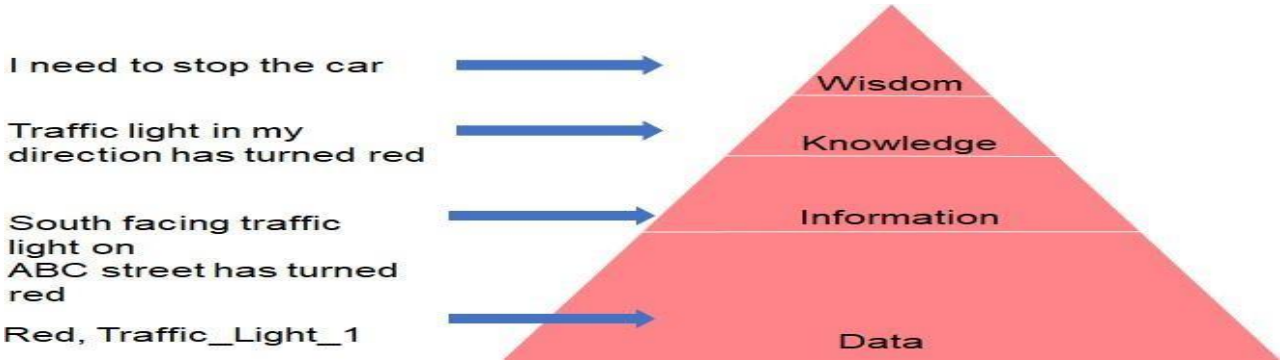
Data Pyramid is made of different stages of working with data

Different Stages of the Data Pyramid

When moving up from the bottom in the Data Pyramid we understand that the pyramid is made up of:

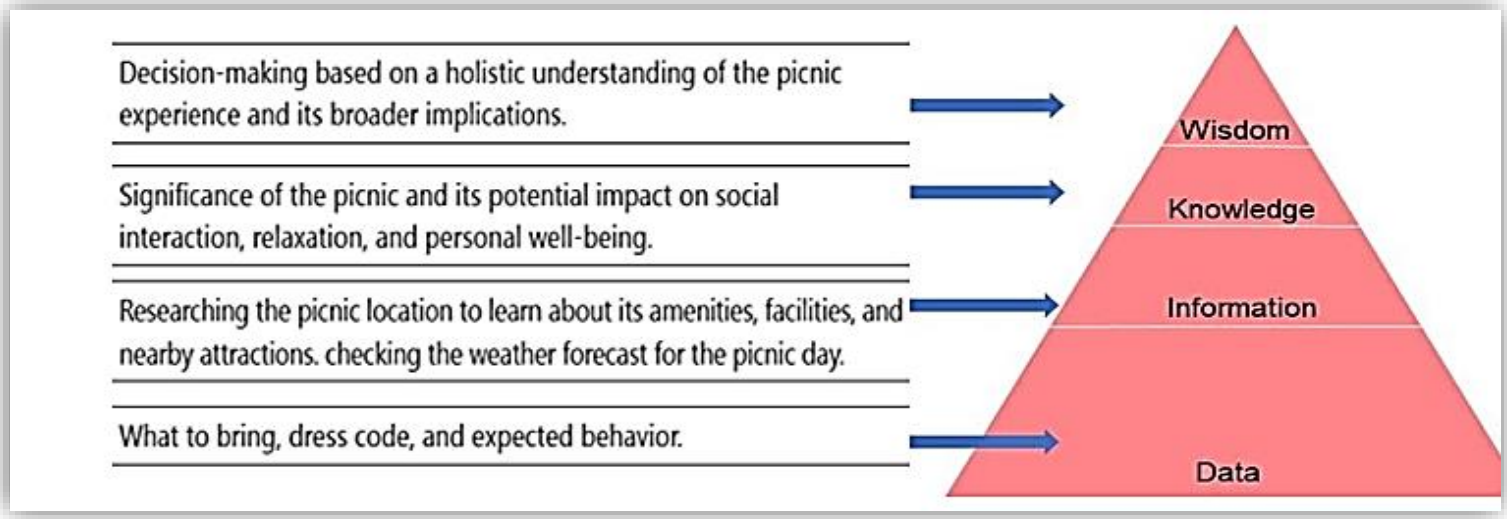
- **Data (Base Level):**
  - In this stage data is in its most basic form, unprocessed and unstructured.
  - It has no meaning and is not very useful in this form.
- **Information:**
  - It is a processed data that collectively carries a logical meaning.
  - It is obtained by analysing raw data to make it useful for decision making.
- **Knowledge:**
  - It is useful information that leads to a deeper understanding.
  - It represents a more profound comprehension of how things happen.
  - It is the ability to use information to achieve desired output.
- **Wisdom (Top Level):**
  - It is the highest level of understanding.
  - It is the ability to understand why things are happening in a particular way.
  - It involves critical thinking to interpret data and make good consistent decisions.

Let's understand Data Pyramid with a simple Traffic Light example:



Let us do another example of creating a data pyramid for planning a picnic as a student after receiving the official letter from school:

Why is Data Literacy Essentials?



1. Informed Decision-Making

- Understanding data helps individuals and organizations make **evidence-based decisions** rather than relying on guesswork or assumptions.
- Example: A business can identify trends in customer behavior and adapt its strategy accordingly.

💼 **2. Workplace Relevance**

- Most jobs today involve some level of data handling—whether it's reading reports, analysing metrics, or making forecasts.
- Data-literate employees are **more efficient**, **more valuable**, and **more competitive** in the job market.

🧠 **3. Critical Thinking and Problem-Solving**

- Data literacy fosters **analytical thinking**, helping people interpret information correctly, identify biases, and avoid misinformation.

🌐 **4. Navigating the Digital World**

- From social media to health apps, we are constantly exposed to data.
- Being data-literate helps users **understand what data is being collected**, how it's used, and how to **protect their privacy**.

🎯 **5. Improved Communication**

- It enables individuals to **present insights clearly** through visuals, reports, or discussions—making collaboration easier and more productive.

🚀 **6. Driving Innovation**

- Organizations use data to discover new opportunities, test hypotheses, and develop products or services.
- A data-literate culture promotes **experimentation and innovation**.

📊 *Impact of Data Literacy*

Data literacy has an immense impact on various aspects of society like business, education, healthcare, and public policy as given below:

- *Business: It improves the decision-making skills of a person. Data-literate employees can effectively analyse data to gain insights into market trends, customer behaviours, and operational performance.*
- *Education: It empowers the teaching-learning process. Students can engage more deeply with course material, particularly in STEM fields.*
- *Healthcare: Healthcare professionals can use data to improve diagnostics, treatment plans, and patient monitoring. Hospitals and clinics can use data to optimize resource allocation, reduce waste, and improve operational efficiency.*

- *Public policy: Policymakers can use data to design, implement, and evaluate policies more effectively. Data literacy promotes transparency, allowing the public to hold policymakers accountable through data-driven evidence.*
- *Social equity: Data literacy can highlight disparities in areas such as education, healthcare, and employment and can promote social equity. It helps ensure that resources are distributed effectively to areas of greatest need.*

How to become Data Literate?

Data Literate is a person who can interact with data to understand the world around them. Let's understand it with following example:

Scenario: Buying a Video game online

Data literacy helps people research about products while shopping over the internet

How do you decide the following things when we are shopping o

- Which is the cheapest product available?
- Which product is liked by the users the most?
- Does a particular product meet all the requirements?

A data literate person can –

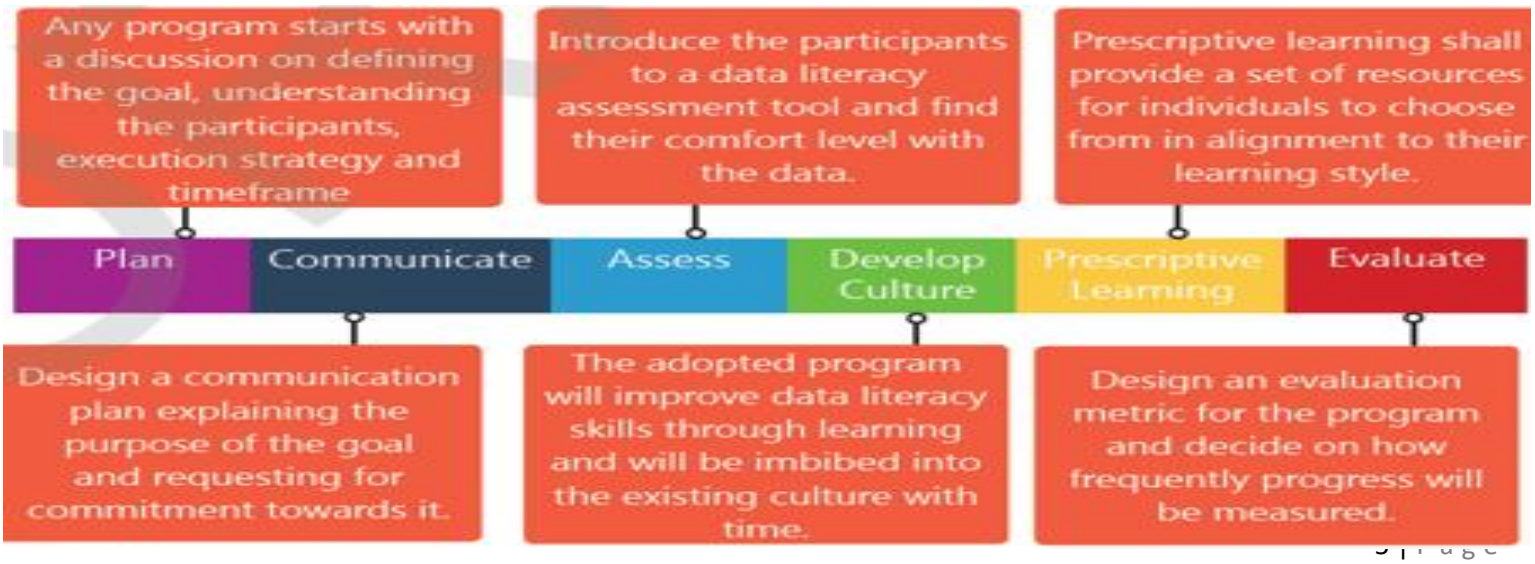
- Filter the category as per the requirement – If the budget is low, select the price filter as low to high
- Check the user ratings of the products
- Check for specific requirements in the product

Data Literacy Process Framework

📘 Data Literacy Process Framework

The data literacy framework provides a comprehensive and structured approach to developing the necessary skills for using data efficiently and with all levels of awareness. Each level builds upon the previous one, fostering a deeper and more understanding ability to work with data.

Here are the typical levels of awareness in a Data Literacy Process Framework:



Plan	✓ Any program starts with a discussion on defining the goal, ✓ understanding the participants, ✓ execution strategy and timeframe.
Communicate	✓ Design a communication plan ✓ explaining the purpose of the goal and ✓ requesting for commitment towards it.
Assess	✓ Introduce the participants to a data literacy assessment tool and ✓ find their comfort level with the data.
Develop Culture	✓ The adopted program will improve data literacy skills through learning and will be imbibed into the existing culture with time.
Prescriptive Learning	✓ Prescriptive learning shall provide a set of resources for individuals to choose from in alignment to their learning style.
Evaluate	Design an evaluation metric for the program and decide on how frequently progress will be measured.

Data Literacy Framework – an Iterative Process

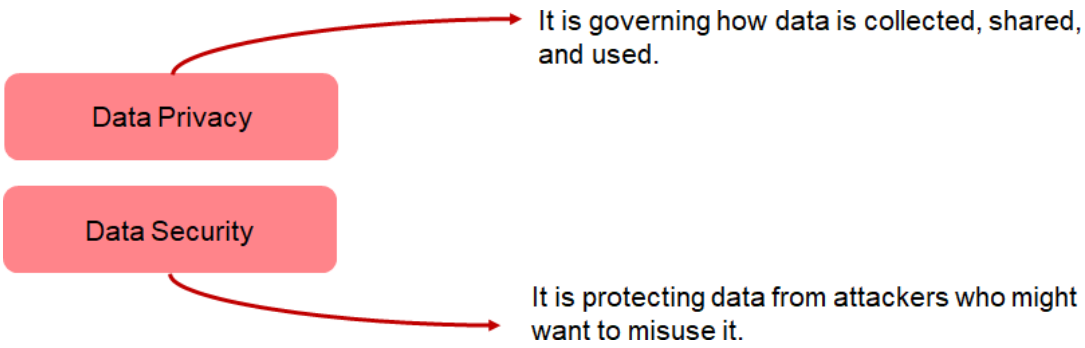
This means the development and enhancement of data literacy skills are not static or one-time events. Instead, they evolve through continuous cycles of learning, application, and refinement.

- **Learning**
  - Learning is the initial stage where individuals acquire new knowledge and skills related to data literacy.
  - Individuals engage in various learning activities such as formal training sessions, online courses, reading materials, and hands-on workshops to gain insights into data concepts, tools, and methodologies.
- **Application**
  - Application involves putting acquired knowledge and skills into practice in real-world contexts.
  - Individuals apply what they have learned to analyse real datasets, solve data-related problems, and make informed decisions.
  - They are engaged in data projects, experiments, or simulations to gain practical experience and develop a deeper understanding of data concepts.
- **Refinement**
  - Refinement focuses on reflecting on past experiences, identifying areas for improvement, and enhancing data literacy skills over time.
  - Feedback from peers, mentors, supervisors, and outcomes of data-related activities informs the refinement process, guiding individuals to adjust their practices accordingly.

Data Security and Privacy

What are Data Security and Privacy? How are they related to AI?

Data Privacy and Data Security are often used interchangeably but they are different from each other. Data privacy determines who can access the data, while data security involves tools and policies to restrict access to the data.

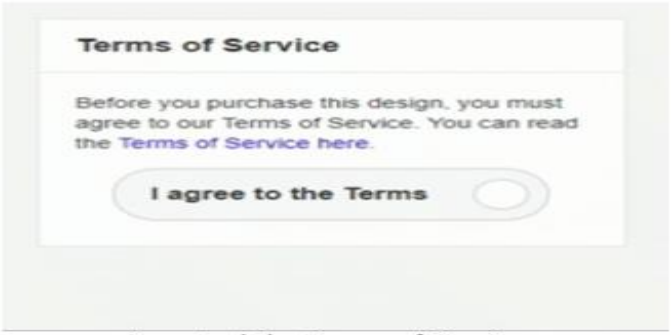


What is Data Privacy?

Data privacy referred to as information privacy is concerned with the proper handling of sensitive data including personal data and other confidential data, such as certain financial data and intellectual property data, to meet regulatory requirements as well as protecting the confidentiality and immutability of the data.

Here are examples of two things which may compromise our data privacy

Downloaded an unverified mobile application



Accepted the Terms of Service without reading

Downloaded an unverified mobile application

A breach at a corporation can put proprietary data in the hands of a competitor.

A data breach at a government agency can put top secret information in the hands of an enemy state.

Why is it important?

The following best practices can help you ensure data privacy:

- Understanding what data, you have collected, how it is handled, and where it is stored.
- Necessary data required for a project should only be collected.
- User consent while data collection must be of utmost importance.

The following best practices can help you ensure data privacy:

- Understanding what data, you have collected, how it is handled, and where it is stored.
- Necessary data required for a project should only be collected.
- User consent while data collection must be of utmost importance.

What is Data Security?

Data security is the practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.

Why Data Security is important?

Due to the rising amount of data in the cloud there is an increased risk of cyber threats. The most appropriate step for such an amount of traffic being generated is how we control and protect the transfer of sensitive or personal information at every known place.

The most possible reasons why data security is more important now are:

- Cyber-attacks affect all the people
- The fast-technological changes will boom cyber attacks

Types of Data Security Controls

Types of Data Security Controls

Different types of data security controls are as follows:

- **Strong Passwords:** Strong password is a combination of upper and lower-case letters, numbers, and special characters that is difficult for unauthorised individuals or automated programs to guess or crack. It is a very basic step that one should take and never share the same with even the most trusted. Avoid using birth dates, anniversary dates, common combinations of numbers.

Some examples of strong passwords are: m#P52s@ap$V, "N4&vQ2! p".

- **Authentication:** It also refers to multi-factor authentication (MFA) and is an additional security layer in online data systems. After a user enters their password to log in, MFA requires them to provide one or more additional forms of authentication to verify their identity. This could include one-time generated code as a security token in smartphones or emails or a fingerprint or facial recognition.

- **Access Controls:** Access controls refer to the security measures and protocols to restrict access to sensitive data, ensuring that only authorised individuals or entities can view, modify, or interact with it. This reduces the risk of unauthorised access by limiting the number of users who can interact with sensitive data.

- **Data Backup:** Data backup refers to the process of creating copies of data to ensure that it can be restored in the event of data loss due to natural disasters, accidents, cyber-attacks, or other unexpected events. Sometimes physical backup media is used to secure in access-controlled environments. Another method to secure data can be on the cloud backup which is considered more reliable.

- **Encryption:** Encryption is a security technique that transforms readable data (plaintext) into an unreadable format (ciphertext) using an algorithm and an encryption key. This process ensures that only authorised individuals with the correct decryption key can access the original data.

- **Data Disposal:** Data disposal refers to the process of securely destroying or deleting data that is no longer needed to prevent unauthorised access, recovery, and misuse. Proper data disposal practices are essential to ensure that sensitive and confidential information does not fall into the wrong hands. Paper documents, CDs, DVDs, and other physical media can be shredded to render them unreadable.

- **Firewall and Antivirus Software:** Using firewall and antivirus software can stop and alert users of any suspicious activity happening on their device. With the timely updated versions of the same, can go a long way in ensuring data security.

- **Training:** Corporates must take up regular Data Security sessions of their staff to sensitise them about following the data protection processes being implemented and the importance of doing so. Making them conscious of suspicious emails, links that they might receive, not leaving their devices unlocked when unattended, keeping software's up to date and not sharing passwords, are some of the things that can be taken up.

- **Audits and Testing of Security System:** Regular audits and testing of security policies, integrated malware protection, firewalls, Wi-Fi connections security, checking applications security, email security and compliance also play very important role in maintaining data privacy and providing data security.
- **Other Basic Preventions:** Being aware of surroundings and threats from insiders, complying with security regulations which might be shared by entrusted agencies or bodies which track online cyber activities all across the world are few other ways to provide cyber security.

Differences between Data Security and Data Privacy

Data Privacy	Data Security
Data privacy ensures the ethical and lawful use of data.	Data security ensures the protection of data from unauthorised access and breaches
It focuses on how data is collected, used, shared, and stored so that the rights of individuals over their data is protected.	It focuses on safeguarding personal data, business data, intellectual property, and many more from various threats.

How are Data Security and Data Privacy related to AI?

Data Security and Data Privacy are crucial components in Artificial Intelligence (AI).

Data Security in AI

AI systems often rely on vast amounts of data for training and operation. Unauthorised access and tampering could lead to inaccurate AI models and compromised outcomes. Many AI applications process sensitive data, such as personal, financial, or health-related information. Strong data security measures can stop data breaches and unauthorised access.

Data Privacy in AI

Data privacy brings the ethical use of AI. This ensures that AI systems comply with data privacy laws and regulations (such as GDPR, CCPA) to help protect individuals' rights and maintain public trust. AI systems must ensure that data is collected and used in ways that users have explicitly consented to, maintaining transparency and trust.

The following are the list of **Do's** and **Don't 's** to be followed for best practices of cyber security:

Do's	Don't's
• Use strong, unique passwords with a mix of characters for each account. • Activate Two-Factor Authentication (2FA) for added security. • Download software from trusted sources and scan files before opening. • Prioritize websites with "https://" for secure logins. • Keep your browser, OS, and antivirus updated. • Adjust social media privacy settings for limited visibility to close contacts. • Always lock your screen when away. • Connect only with trusted individuals online. • Use secure Wi-Fi networks. • Report online bullying to a trusted adult immediately.	• Avoid sharing personal info like real name or phone number. • Don't send pictures to strangers or post them on social media. • Don't open emails or attachments from unknown sources. • Ignore suspicious requests for personal info like bank account details. • Keep passwords and security questions private. • Don't copy copyrighted software without permission. • Avoid cyberbullying or using offensive language online.

YouTube -- easy2learn with jyoti Website—distancediary.com

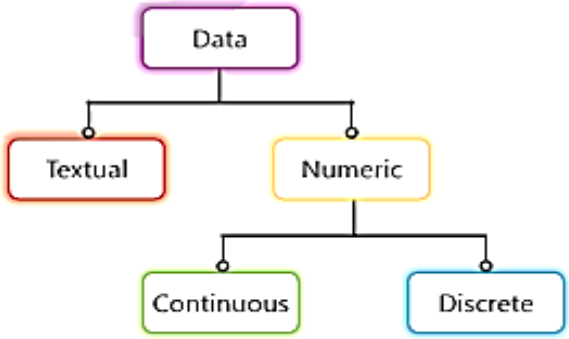
Acquiring Data, Processing, and Interpreting Data

Working with data involves three key steps: acquiring, processing, and interpreting. First, gather data from sources like surveys and databases. Next, process it by cleaning and organising it for accuracy.

Finally, analyse the data to find patterns and insights that help make informed decisions.

Types of Data

In statistics, various types of data are gathered, analysed, interpreted, and presented. These data consist of individual factual pieces recorded for analysis. Data analysis involves interpretation and presentation, producing statistics. Data classification and handling are crucial processes that use multiple tags and labels to define data, ensuring its integrity and confidentiality. Artificial Intelligence is crucial, with data serving as its foundation. We come across different types of data and information every day.



Data can be broadly classified under Textual data and Numeric Data as explained.

Textual Data (Qualitative Data)

Textual data is the information that is written or expressed using words and language. It includes things like articles, books, emails, messages, and any other written content. Instead of numbers, it's made up of letters, words, and sentences that convey meaning and information.

Example: "Learning AI is fun"

Numeric Data (Quantitative Data)

Numerical data means information that's in numbers, not words or descriptions. It's often called quantitative data because it's collected as numbers and can be used for math and stats. For instance, if you know the total number of workers and how many are men, you can figure out how many are women by subtracting. This ability to do math with numerical data makes it great for doing statistics and analysing data.

For example, Marks, Temperature, etc.

Numeric data can be further classified as Continuous Data and Discrete Data:

Continuous Data	Discrete Data
Continuous data can take as a numeric value given within a range.	Discrete data refers to countable, distinct values. It consists of whole numbers without decimal parts that represent distinct categories or values.
This type of data can be infinitely subdivided and often includes decimal points.	Discrete data cannot be subdivided meaningfully.
Often used to analyse using statistical techniques such as mean, median, standard deviation, and correlation.	It is used to analyse using frequency distributions, bar charts, and probability distributions.
Example: Dimensions of classroom, Height, Weight, Temperature, Time, etc.	Examples: Number of girls and boys in class, Number of subjects in class 9th, Count of anything.

AI Domains and Type of Data

Various types of data are utilised across different domains to train models, make predictions, and generate insights. Here are the types of data commonly used in three key domains of AI.

Natural Language Processing (NLP)

Natural Language Processing (NLP) is all about teaching computers to understand and work with human language. Types of data used in NLP is:

- **Textual Data**: This includes a wide range of written text, such as articles, books, emails, social media posts, web content, PDF files, etc.
- **Audio Data:** Audio recordings of spoken language, which are transcribed into textual data.

Computer Vision

Computer Vision is like giving eyes to computers. It helps them look at pictures and videos from the real world and understand what they're seeing. With Computer Vision, computers can figure out what's in a picture or video, just like we do. They can recognize objects, people, and even actions happening in videos.

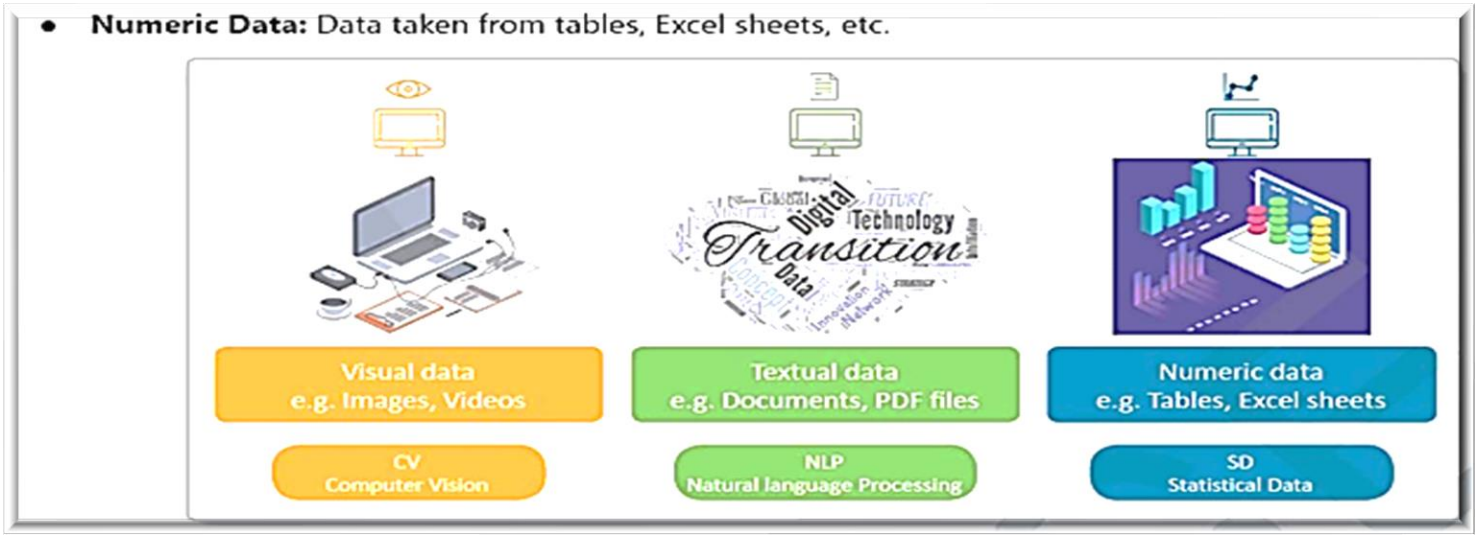
Types of data used in Computer Vision include:

- **Image Data**: Digital images captured by cameras or satellite imagery, medical scans, and surveillance footage.
- **Video Data**: Video data captured using camera

Machine Learning

Machine Learning is like teaching computers to learn from examples and make decisions on their own. Imagine if you showed a computer lots of pictures of dogs and cats, and you told it which ones were dogs and which ones were cats. After seeing many examples, the computer learns to tell dogs and cats apart on its own. Types of data used in Machine Learning include:

- **Numeric Data:** Data taken from tables, Excel sheets, etc.



Ai Task

Let us now do an exercise to categorise the given data as Textual Data (Qualitative Data) or Numeric Data (Quantitative Data):

Temperature

Gender

Show size

Comment on social media

Favourite colour

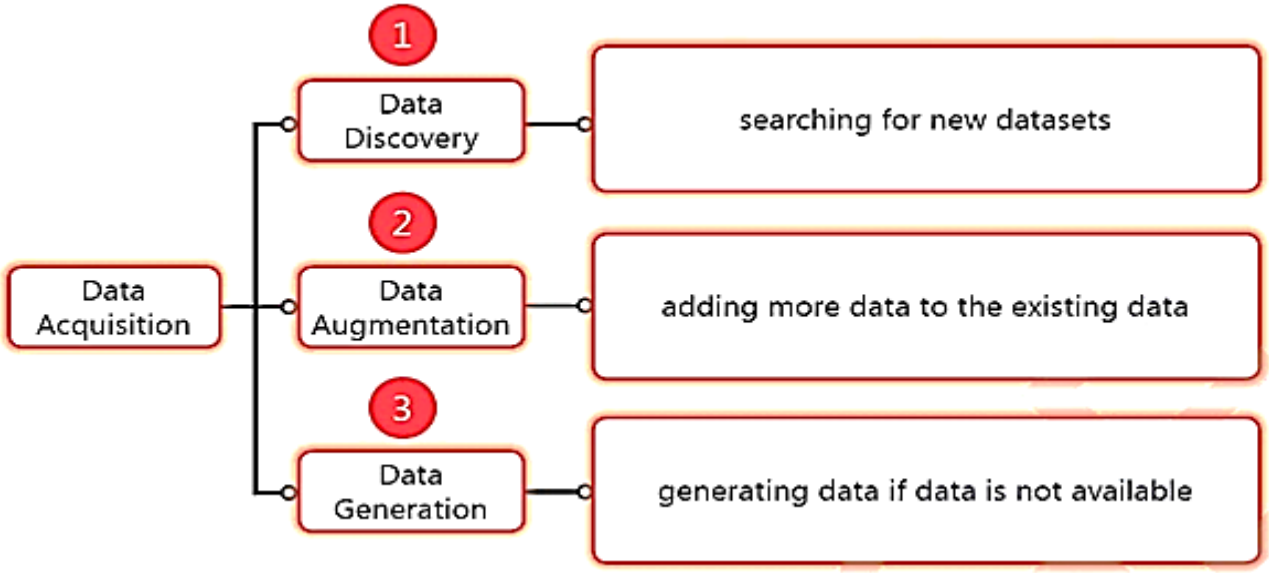
Newspaper article

Population number in a state

Email

Data Acquisition/Acquiring Data

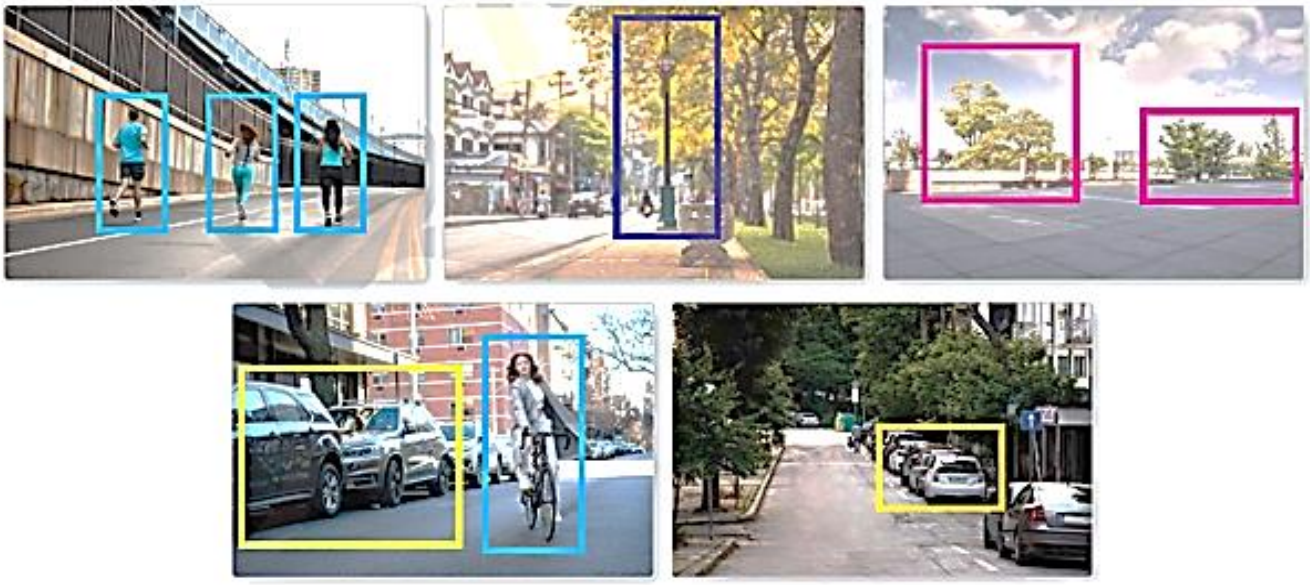
Data acquisition, also known as acquiring data, refers to the procedure of gathering data like raw facts, figures or statistics from relevant sources either for reference or for analysis needed in AI projects. This involves searching for datasets suitable for training AI models. The process typically comprises three key steps and plays a crucial role in obtaining and preparing data for analysis.



Step 1: Data Discovery

Data discovery is about hunting for valuable information in different places, checking if it's good quality, and making sense of what we find. In the above example:

- We will require pictures of roads and the objects on roads.
- We can search and download this data from the Internet.



Step 2: Data Augmentation

Data augmentation is the process of increasing the amount and diversity of data. We do not collect new data, rather we transform the already present data. Data augmentation means increasing the amount of data by adding copies of existing data with small changes. The image given here does not change, but we get data on the image by changing different parameters like colour, rotation, flipping and brightness. New data is added by slightly changing the existing data.

In the above example:

- We apply flipping and rotation transformation to create variations of the original images.
- We also simulate occlusions such as objects partially blocking the view to train the model to handle obstructed scenarios.

Step 3: Data Generation

Data generation refers to generating or recording data using sensors. Recording temperature readings of a building is an example of data generation. Recorded data is stored in a computer in a suitable form.

In the above example, creating fake driving scenarios with different road conditions, traffic patterns, weather, and lighting to cover many possible situations.

Sources of Data Acquisition

Various Sources for Acquiring Data can be Primary or Secondary as given below:

Primary Data Sources

The data generated from the experiment is an example of primary data. Some of the sources for primary data include surveys, interviews, experiments, etc. Here is an Excel sheet showing the data collected for students of a class.

Name	Gender	Address	Phone	Percentage
Ridham	Female	A-12 Ramesh Nagar	5555555555	92
John	Male	K-15 Pitampura	5557777888	89
Rohit	Male	G-2 Faridabad	2323232323	90
Gaurav	Male	L-3 Karolbaogh	3987678931	87.5

Secondary Data Sources

Secondary data sources are the external sources for collecting data, rather than generating it personally. Some sources for secondary data collection are: Published Literature, Government Publications, Market research reports, etc.

Best Practices for Acquiring Data

1. **Set Clear Goals:** Understand why you need the data and what you want to achieve; specify the type, format, and detail level required.
2. **Identify Data Sources:** Use primary data you collect yourself (surveys, interviews) and secondary data from others (reports, databases).
3. **Evaluate Sources:** Ensure data sources are trustworthy, relevant, accurate, and current; get necessary permissions and respect privacy.
4. **Collect and Prepare Data:** Use surveys, interviews, sensors, and web scraping; clean data by fixing errors, removing duplicates, and anonymising.
5. **Validate, Document, and Store:** Cross-check and sample for accuracy, keep detailed records and meta-data, store data securely, and regularly update it while following laws and regulations.

Checklist of factors that make data good or bad

Here's a checklist of factors that can help determine whether data is of good quality (good data) or poor quality (bad data):

Good Data

- Information is well structured
- It is accurate
- It is consistent
- It is cleanly presented
- Contains information which is relevant to our requirement

Bad Data

- Information is scattered
- Contains a lot of incorrect values
- Contains missing and duplicate values
- It is poorly presented
- Contains information which is not relevant to our requirement

Data Acquisition from Websites

The process of collecting data from websites using software is called Data Scraping. It is a common method for extracting information from websites.

Just like you might copy text from a book or form your friend, data scraping involves copying information from websites. But instead of doing it manually, we use special tools or programs to do it automatically. These tools can navigate websites, find the information we want, and copy it into a format we can use.

We scrape websites to get data we need for different reasons. Maybe we're collecting prices for market research, news articles for analysis, or customer reviews for a product.

While web scraping is not illegal, using data without permission is illegal. Think of web scraping like picking fruit from someone else's garden without their permission and it is also about what you do with the fruit afterwards

Using data with permission is legal and ethical, just like getting permission to take fruit. It's all about respecting the rights of the website owner and following the rules.

Ethical Concerns in Data Acquisition

While gathering data and choosing datasets, certain ethical issues can be addressed before they occur:

Bias	Take steps to understand and avoid any preferences or partiality in data
Consent	Take necessary permissions before collecting or using an individual's data
Transparency	Explain how you intend to use the collected data and do not hide intentions
Anonymity	Protect the identity of the person who is the source of data
Accountability	Take responsibility for your actions in case of misuse of data

Usability of Data

Imagine completing a school project. You need clear instructions, a neat workspace, and accurate information. Similarly, using data effectively relies on its clarity, organisation, and accuracy. There are three primary factors determining the usability of data:

1. **Structure of Data:** Defines how data is stored. Data needs to have a clear structure. It should be organised in a way that makes sense so that it can be used effectively.

   For example:

   Marks of a students arranged in a spreadsheet.

Student ID	Class	Section	Name	Percentage
10187	12	D	Rohit Rawat	72%
10013	12	B	Ashish Gupta	85%
10143	12	C	Vishal Garg	65%
11919	11	E	Chandan Bhatia	89%
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65%
11518	11	B	Deepak Vashisht	81%
11213	11	A	Deepti Verma	95%
10311	10	C	Vashali Gurung	93%
11095	11	A	Misha Malhotra	83%

Spreadsheet – Good structure
Data is stored in a sheet with the details of each individual stored according to a set of rules.

Rohit Rawat a student with ID 10187 of Class 12 Section D has scored 72%.

Text document – Poor structure
Data is stored in a text document with no set of organising rules.

When your mother starts cooking your favourite food she ensures before cooking that all ingredients are available and are put in order for smooth and organised cooking.

2. **Cleanliness:** Clean data should not have duplicates, missing values, outliers, and other anomalies so that its reliability and usefulness for analysis is not affected. In this particular example, cleaning of data removes the duplicate values.

Student ID	Class	Section	Name	Percentage
10187	12	D	Rohit Rawat	72
10013	12	B	Ashish Gupta	85%
10143	12	C	Vishal Garg	65%
11919	11	E	Chandan Bhatia	89
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65
11518	11	B	Deepak Vashisht	81%
11213	11	A	Deepti Verma	95
10311	10	C	Vashali Gurung	93%
11095	11	A	Misha Malhotra	83%

Student ID	Class	Section	Name	Percentage
10311	10	C	Vashali Gurung	93%
11213	11	A	Deepti Verma	95%
11095	11	A	Misha Malhotra	83%
11518	11	B	Deepak Vashisht	81%
11919	11	E	Chandan Bhatia	89%
10013	12	B	Ashish Gupta	85%
10578	12	C	Ruchi Sharma	91%
10143	12	C	Vishal Garg	65%
10187	12	D	Rohit Rawat	72%

3. **Accuracy:** Accuracy is same as reliability so it indicates how well the data matches real-world values. Accurate data closely reflects actual values without errors, enhancing the quality and trustworthiness of the dataset.

   When your measurement is accurate, it makes your data really good. It's like having a gold star on your homework—it shows you did a great job!

   In the example given below, we are comparing data gathered from measuring the length of a small box in centimeters.

Weight of 12 Eggs in a Box (gms)	Average
1116	
840	
1080	720
1200	
720	
Inaccurate Data	Average

Weight of 12 Eggs in a Box (gms)	Average
600	
840	
660	720
780	
720	
Accurate Data	Average

YouTube -- easy2learn with jyoti Website—distancediary.com

Features of Data

Data features are also called the characteristics or properties of the data. They describe each piece of information in a dataset. They define what each data point represents and help us make sense of the data. For example,

- In a table of student records, features could include things like the student's name, age, or grade.
- In a photo dataset, features might include properties like the colour present in each image, the resolution, brightness, or the presence of certain objects.

These features help us understand and analyse the data. In AI models, we need two types of features: Independent and Dependent.

Independent

Independent variables (sometimes called predictor variables) are those that are used to generate predictions about or to account for the variation in the dependent variable (the goal). These features are the input to the model—they're the information we provide to make predictions.

Dependent

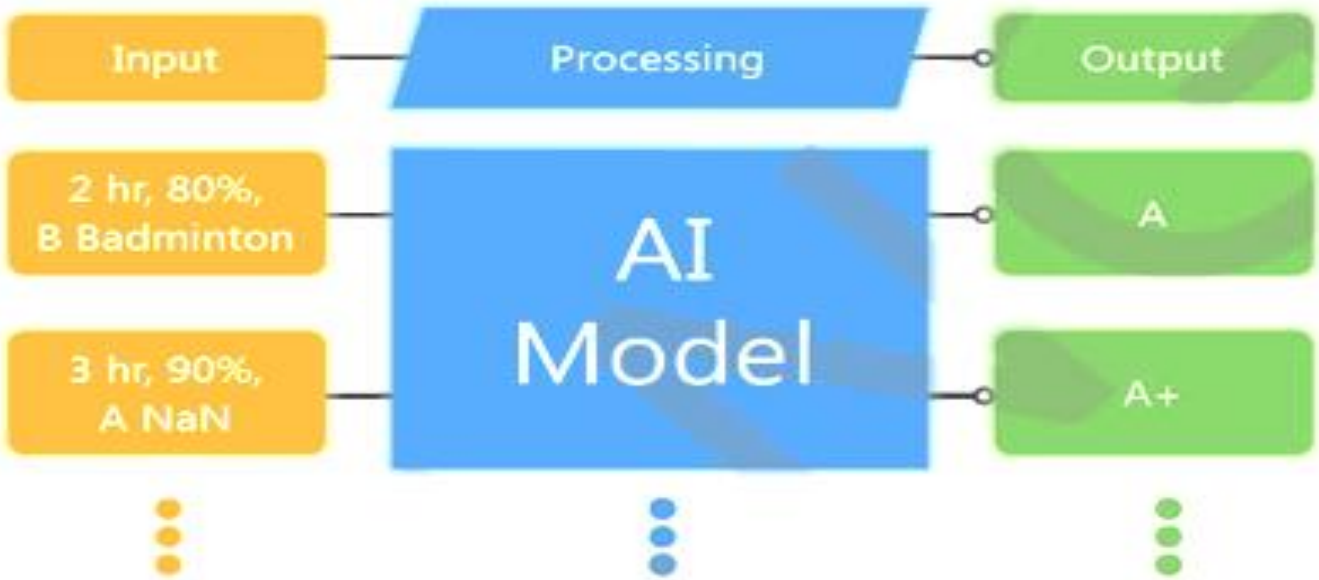
The dependent variable is the variable about which predictions or explanations are being sought. These features are the outputs or results of the model—they're what we're trying to predict. For example, imagine we're building an AI model to predict students' final exam grades based on various factors. The independent features would include:

- **Study Hours:** The number of hours a student spends studying.
- **Attendance:** Whether the student attended classes regularly or not.
- **Previous Grades:** The grades the student received in previous exams.
- **Extracurricular Activities:** Participation in extracurricular activities, such as sports or clubs.

The dependent feature, in this case, would be:

- the final exam grade—the outcome or prediction that the model gives us.

Together, they help us understand and improve student outcomes using AI-driven predictions.



Data Preprocessing

Data preprocessing is an essential phase in the machine learning process that prepares datasets for effective machine learning applications. It includes multiple processes to clean, transform, reduce, integrate, and normalise data.

YouTube -- easy2learn with jyoti Website—distancediary.com

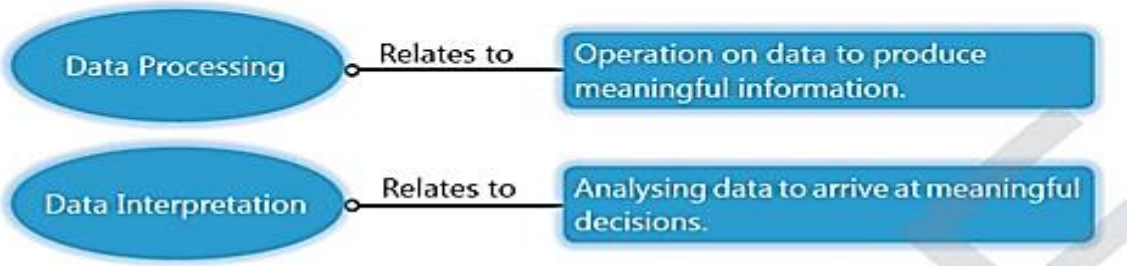
Data Processing and Data Interpretation

Data processing means preparing and analysing raw information to train models or predict outcomes, including tasks like cleaning and training. Data interpretation in AI involves analysing model outputs to understand patterns, refine models, and make informed decisions.

Observe and answer the following:

- How many large lollipops are there?
- If each large lollipop represents 5 units of sweetness, how much total sweetness do the three lollipops represent?

- Among the small round candies, which colour appears most frequently?
- What is the ratio of green round candies to blue round candies?
- How many pink round candies are there in the image?



Data Processing

Data processing involves tasks to refine raw data for analysis or application, including cleaning, organising, transforming, and summarising information.

- It ensures data accuracy, relevance, and accessibility for effective decision-making and analysis.
- It is crucial across various sectors like business, science, and technology, facilitating better utilisation of data assets.
- Data processing helps computers understand raw data.
- Use of computers to perform different operations on data is included under data processing.

Data Interpretation

Data interpretation is the process of making sense of data by analysing it to uncover patterns, trends, and insights. It involves examining the data to understand its meaning, implications, and significance, helping to inform decision-making and draw conclusions.

- It is the process of making sense out of data that has been processed.
- The interpretation of data helps us answer critical questions using data.

YouTube -- easy2learn with jyoti Website—distancediary.com

Process of Data Interpretation

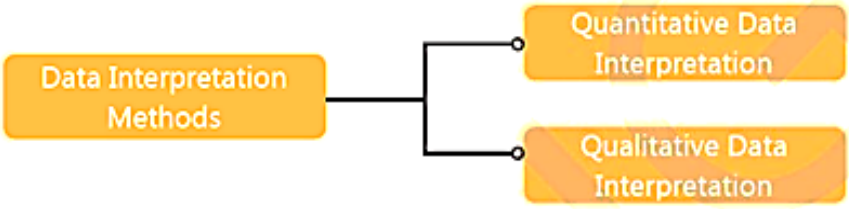
- **Acquire:** This initial step involves gathering raw data from diverse sources such as surveys, databases, or sensors. It ensures that all relevant information is collected to provide a comprehensive dataset for analysis.
- **Process:** Once the data is collected, it undergoes cleaning and organisation to remove errors, inconsistencies, or irrelevant information. This step ensures that the data is in a standardised format and ready for further analysis.
- **Analyse:** In this phase, the cleaned and organised data is scrutinised to identify patterns, correlations, or trends. Statistical methods, algorithms, or data visualisation techniques may be employed to extract meaningful insights from the data.
- **Interpret:** After analysing the data, the results are interpreted to derive actionable insights or conclusions. This involves understanding the implications of the analysis findings in the context of the problem or question at hand.
- **Present:** The final step involves presenting the interpreted findings in a clear and engaging manner. This could include visualisations such as graphs or charts, along with concise summaries, to effectively communicate the insights derived from the data analysis.

These steps make sure that working with data is organised, complete, and useful, so that organisations can make smart choices based on the data.

Methods of Data Interpretation

Data Interpretation is the process of making sense out of a collection of data that has been processed. This collection may be present in various forms like bar graphs, line charts and tabular forms and other similar forms.

There are two ways to interpret data-



Quantitative Data Interpretation

It is the process of analysing and understanding numeric data. This type of data often comes from surveys, experiments, and numerical measurements. Quantitative data provides statistical insights and helps in identifying patterns and trends.

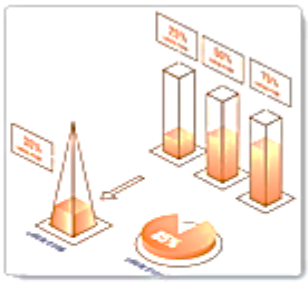
- The interpretation of quantitative data focuses on measurable outcomes and numerical relationships.
- It helps us answer questions like "when," "how many," and "how often".

For example: (how many) numbers of likes on the Instagram post.

Data Collection Methods in Quantitative Data Interpretation

Data collection methods in quantitative data interpretation involve systematic techniques like surveys and experiments to gather numerical data. These approaches ensure data accuracy, facilitating reliable analysis and inference across various fields such as social sciences and healthcare.

YouTube -- easy2learn with jyoti Website—distancediary.com


- **Interviews:** Quantitative interviews play a key role in collecting information.
- **Polls**: A poll is a type of survey that asks simple questions to respondents. Polls are usually limited to one question.

- **Observations**: Quantitative data can be collected through observations in a particular time period.
- **Longitudinal Studies**: A type of study conducted over a long time
- **Survey**: Surveys can be conducted for a large number of people to collect quantitative data.

Steps to Quantitative Data Analysis

The four steps involved in quantitative data analysis are:

1. **Relate measurement scales with variables:** Understand the type of data you have and match it with the appropriate measurement scale. Identify the variables you are studying (e.g., height, age, income). For example, if you are looking at student grades, use ordinal (A, B, C) or ratio (numeric scores) scales.
2. **Connect descriptive statistics with data:** Use statistical measures to summarise and describe your data.

   For example: If you have test scores for a class of students, calculate the mean score to know the average performance, the median to understand the central tendency, and the standard deviation to see how varied the scores are.
3. **Decide a measurement scale:** Choose the appropriate scale to measure your data based on the type of variable and the level of detail required. For example, for measuring weight, use a ratio scale because weight can be zero and can be measured precisely.
4. **Represent data in an appropriate format:** Display your data in a way that makes it easy to understand and interpret. For example, if you want to show the distribution of test scores in a class, you might use a histogram. If you want to show the relationship between study hours and test scores, a scatter plot would be appropriate.

Qualitative Data Interpretation

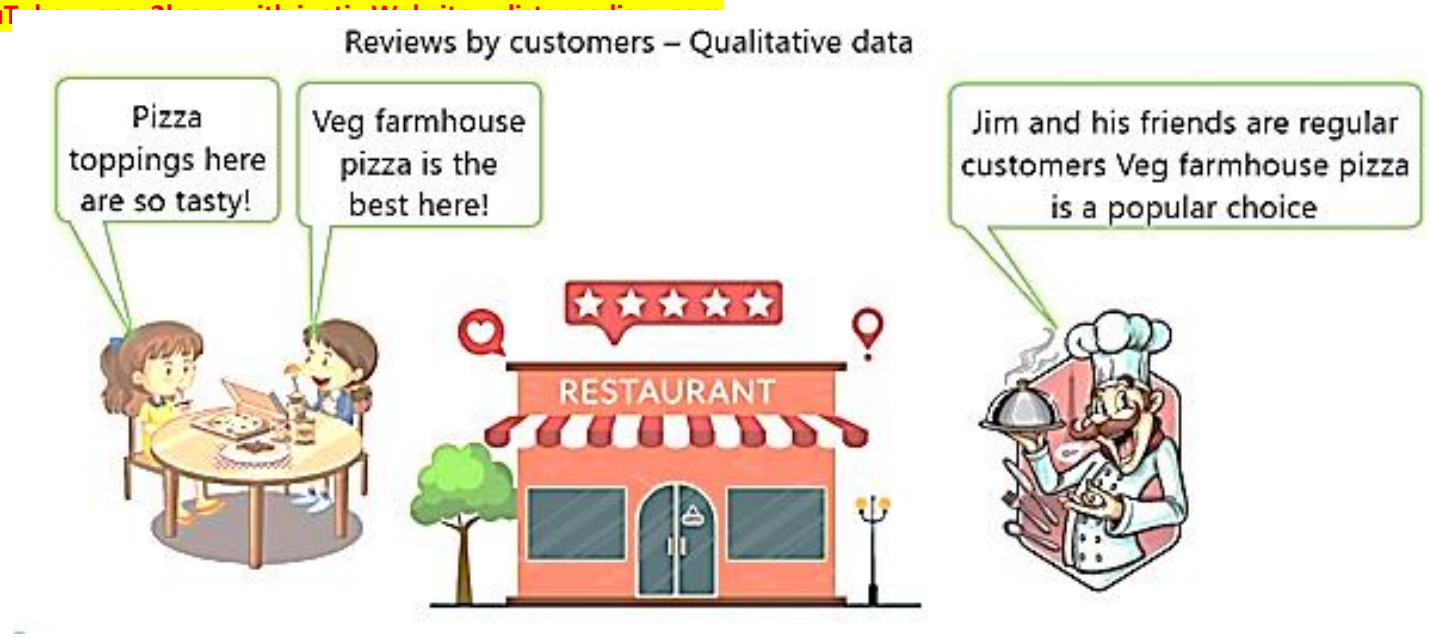
It is the process of analysing and understanding non-numeric data. This type of data often comes from interviews, surveys, observations, or textual content. Qualitative data tells us about the emotions and feelings of people. Qualitative data interpretation is focused on insights and motivations of people

Data Collection Methods in Qualitative Data Interpretation

Data collection methods in qualitative data interpretation involve techniques such as interviews and observations to gather rich, descriptive data for nuanced analysis, fostering deeper understanding of complex human experiences and behaviors. Some methods are as follows:

- **Record keeping**: This method utilises the documents that are reliable and well curated and other similar sources of information as the data source that are verified and maintained. It is similar to going to a library.
- **Observation**: In this method, data is collected by observing the participants, their behavior and emotions, carefully,
- **Case studies**: In this method, data is collected from case studies.
- **Focus groups**: In this method, data is collected after a group discussion on topics of relevance.
- **Longitudinal studies**: In this data collection method, data is collected on the same data source repeatedly over an extended periodof time.
- **One-to-one interviews**: In this method, data is collected using a one-to-one interview.

YouT



Steps to Qualitative Data Analysis

The five steps involved in qualitative data analysis are:

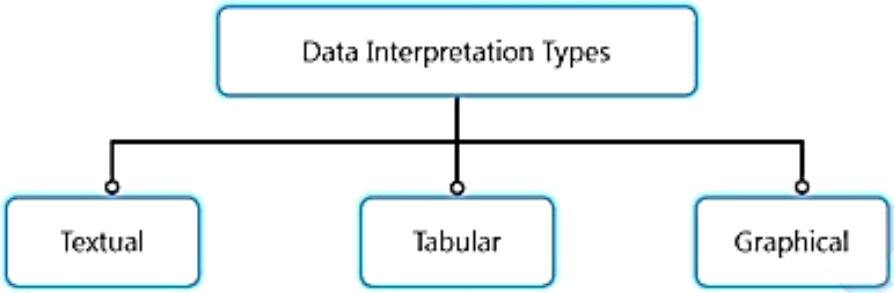
1. **Collect data:** Gather qualitative data through various methods to understand people's experiences, opinions, or behaviors. This is done through interviews, surveys, observations, or documents. For example, a researcher interviews patients about their experiences with a new healthcare app, recording their responses for further analysis.
2. **Organize:** Prepare and arrange the collected data in a systematic way to make it easier to work with. For example, The researcher transcribes the recorded interviews into text documents and organizes them by participant or interview date.
3. **Set a code to the data collected:** Assign labels or codes to different parts of the data to identify themes, patterns, or categories. For example, the researcher reads through the interview transcripts and highlights sections discussing "ease of use," "technical issues," and "benefits of the app," tagging them with corresponding codes.
4. **Analyse your data:** Examine the coded data to identify deeper patterns, relationships, and insights. For example, The researcher groups codes related to "ease of use" and "technical issues" into a broader theme of "user experience" and analyses how these themes impact overall user satisfaction with the app.
5. **Reporting:** Present the findings clearly, using quotes and visual aids to support your conclusions and recommendations. For example, the researcher writes a report highlighting the main themes along with positive and negative feedback.

Difference Between Qualitative and Quantitative Data Interpretation

Qualitative Data Interpretation	Quantitative Data Interpretation
Categorical	Numerical
Provides insights into feelings and emotions	Provides insights into quantity
Answers how and why	Answers when, how many or how often
Methods – Interviews, Focus Groups	Methods – Assessment, Tests, Polls, Surveys
Example question – Why do students like attending online classes?	Example question – How many students like attending online classes?

Types of Data Interpretation

There are three ways in which data can be presented:



Textual DI

Data is put into words, like in a paragraph, which works well for small amounts of data that can be easily understood. But for larger amounts, this type of presentation may not be the best because it can get too complicated. For instance, a paragraph might describe how a company's sales went up in the first quarter, and how many units of each product they sold, as well as improvements in customer satisfaction.

Tabular DI

Data is organised systematically in rows and columns within a table, facilitating structured representation. The title of the table, "Students Marks Analysis," provides a descriptive overview of the table's content, summarising the analysis of student marks within the table.

Table: Students' Marks Analysis

Student Name	English	Maths	Science	Percentage
Rehan	85	90	88	87.67%
Hemant	78	82	80	80.00%
Deepak	92	88	91	90.33%
Lucky	74	76	78	76.00%
Ramesh	88	84	85	85.67%
Saima	90	89	92	90.33%
Abdul	80	78	79	79.00%
Seema	85	87	86	86.00%

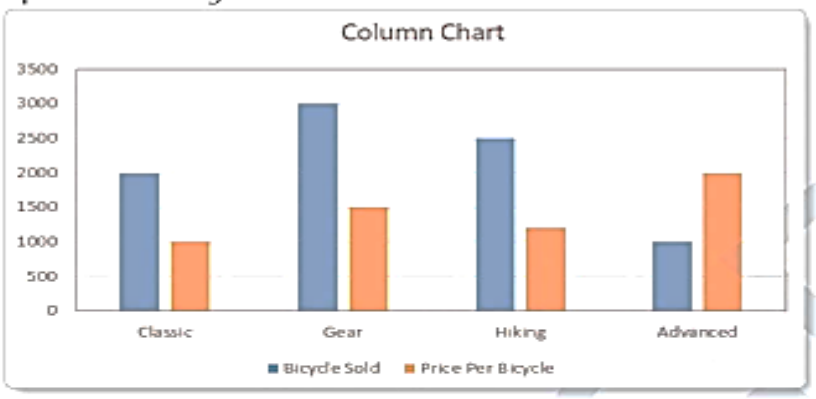
Graphical DI

Some of the graphs include bar graphs, line graphs, pie charts, and scatter plots, which help in visualising trends, relationships, and distributions within the data.

YouTube -- easy2learn with jyoti Website—distancediary.com

Bar Graphs

In a Bar Graph, data is represented using vertical and horizontal bars.

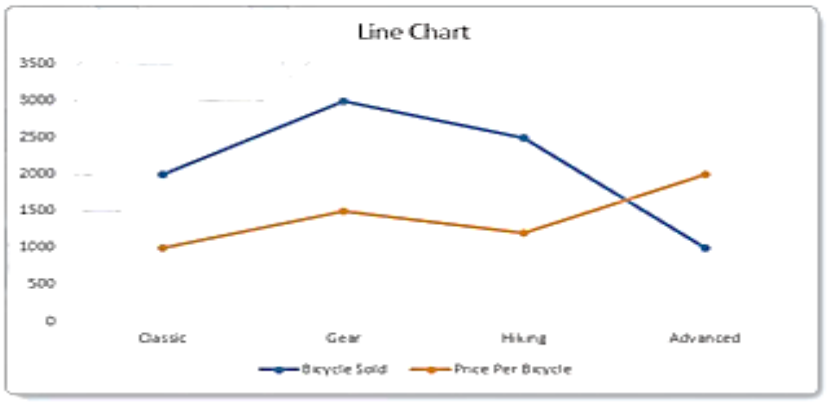


Pie Charts

Pie charts resemble pies, with each slice representing a portion of the whole pie assigned to different categories. These circular charts are divided into sections, and the size of each section corresponds proportionally to its value within the dataset.

Line Graphs

A line graph connects data points to illustrate changes in quantity over time, aiding in visualising trends and patterns.



Importance of Data Interpretation

Data interpretation is crucial as it transforms raw data into actionable insights, guiding informed decision-making. By analysing and understanding data, organisations can uncover trends, patterns, and relationships, enabling them to optimize strategies, mitigate risks, and drive growth.

Informed Decision Making

A decision is only as good as the knowledge it is based on. It means When we analyse data, we get a clearer picture of what's going on. This helps us make decisions that are more likely to lead to success.

For example, if the average height of students is known, school can custom design the chairs and tables according to the requirement of the class.

Reduced Cost

Identifying needs can lead to reduction in cost. It means by knowing what's necessary, we can cut down on waste. We can use resources more efficiently and not spend money on things that aren't important.

For example, restaurant owner could decide to drop/modify some dishes of the menu which aren't popular or have got bad reviews.

Identifying Needs

We can identify the needs of people by data interpretation. It means understanding what people want or require by looking at the information we have.

For example, in a Pizza Shop there are possibilities that Veg Farmhouse Pizza is a popular choice among age group 8-10.

Using Tableau for Data Presentation

Using Tableau for data presentation involves connecting to various data sources, creating diverse visualisations, and enabling interactive features. It supports sharing and collaboration, offers advanced analytics capabilities, and promotes best practices for clear and effective data communication.

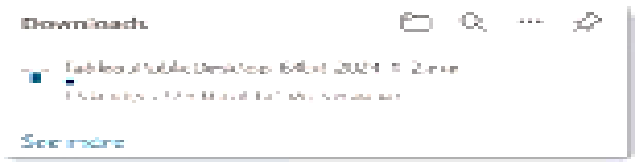
What is Tableau?

Tableau is a powerful data visualisation and business intelligence tool for visualising and analysing data in order to aid in business choices. It takes in data and produces various charts, graphs, maps, dashboards, and stories.

Steps to Download Tableau

1. Visit the link https://public.tableau.com/en-us/s/download
2. Click on **DOWNLOAD TABLEAU PUBLIC.** It will display the given screen where you enter your details.

3. Click on the **DOWNLOAD THE APP** button to begin with the download process as shown below:



4. After finishing with the downloading of the files, double-click the installer of the Tableau Public Desktop. The Tableau Public 2024.1 Setup wizard opens.
5. Select the **I have read and accept the terms of the license agreement** check box to accept the terms of the license agreement.